

フィルタリングの効果って？

ROATは、一定の条件を満たしてアップロードされたHTTPD(Apacheなど)のログをフィルタリングし、エンドユーザの利用と見なされるダウンロードを抽出して分析します。

HTTPDのログには、エンドユーザが文献を利用するためにアクセスしたログの他に、検索エンジンのロボットによるアクセス、システムの応答など、さまざまなログが含まれています。そのままでは、文献がアクセス先になっているログをカウントしても、エンドユーザの利用について知ることはできません。

研究者に「先生の論文、先月は〇回利用されましたよ！」と自信をもって言うためには、フィルタリングが不可欠といえるでしょう。

ログのフィルタリングでは、次のような事項に着目しています。
(詳細は、H19年度CSI報告交流会, ROATワークショップ(2009.2.18)の 佐藤義則東北学院大学教授の資料をご覧ください)

HTTPのステータスコード

HTTPのステータスコードは、サーバの応答結果を示す3桁の数字です。ROATはCOUNTER Code of Practiceに沿って、コードが200と304のログ以外を除去しています。

検索エンジン等のロボットのアクセス

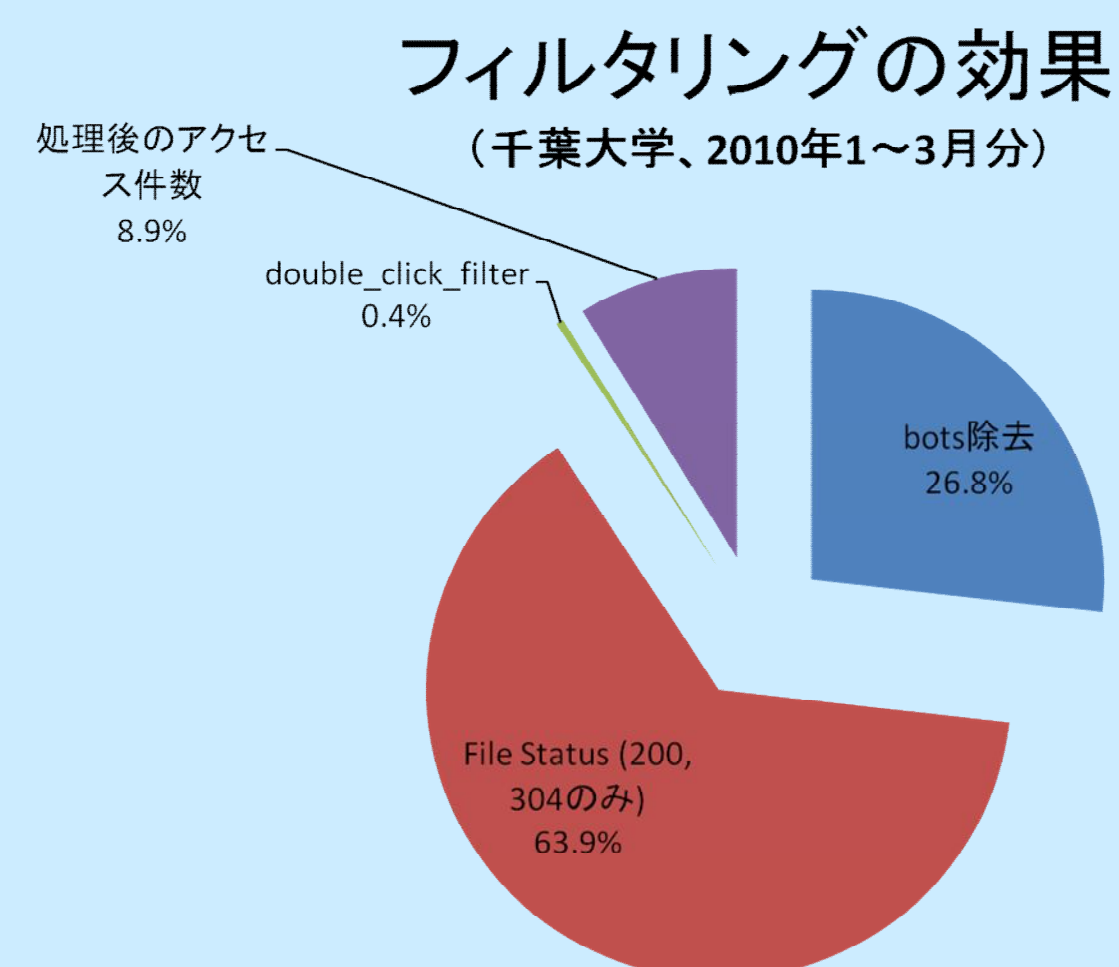
bot, クローラーなどと呼ばれ、多くの種類があります。ROATで管理しているロボットリスト(※)にしたがって除去しています。
どの検索エンジン等のロボットが多くアクセスしているか、そのアクセスはログ全体のどれくらいを占めるかなどは各リポジトリによって異なります。H.19年7-12月の9大学のログの比較分析によると、botの割合は最小12.7%, 最大では72.2%も！

※ロボットリストの管理が、実は大変！

検索エンジン等のロボットの種類は、常に増えています。そのため、ロボットのリストを最新に保つことは大変です。そこで、ROATでは利用する機関の皆さんからロボットの情報を提供いただけるシステムを用意しています。これにより「あやしいアクセス」の情報を収集、確認して、ロボットリストを更新することができます。

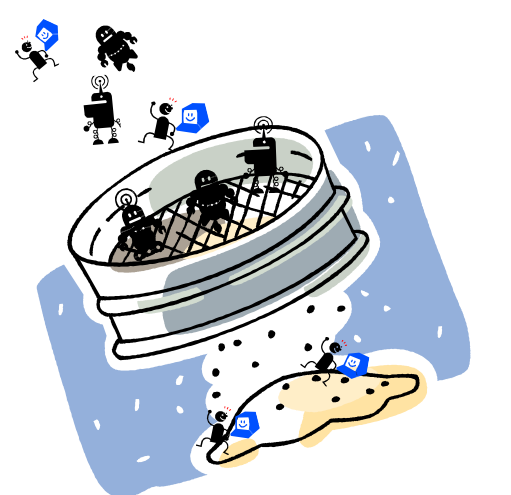
同一IPアドレスからのファイル要求数

COUNTERのCode of Practiceに準拠し、一定時間内に同一IPアドレスから行われたファイル要求(request)のうち最初のを除去します。



ROATは2009-2010でさらに便利になりました！

- **結果ダウンロード時の、書誌事項が充実** → 著者名が第一著者だけから全著者になるなど、項目が増えました。
・JAIROからのデータ取得が毎週に。登録したばかりのデータでも書誌つきでアクセス件数が出力されます。
- **ログの空白期間ができた時の対応が簡単に** → 各機関でアップロード済ログデータを削除できるようになりました。
該当期間を後から埋めることはできないので、空白以降のログを消して、再度アップロードしてください。
- **ログのアップロードミスを少なく** → アップロードしたファイル名が表示され、間違っアップロードしたファイルはその場で削除できます。
アップロードしたファイルの分析は夜間に行われます。




ご利用の希望は… 千葉大学附属図書館 ir@office.chiba-u.jp までご連絡ください！

各国のリポジトリアクセス統計に関する取り組み

- ROAT (Repository Output Assessment Tool, 日本)
COUNTERの標準仕様に基づいたアクセスログの分析を行うシステムを提供。
分析結果に書誌情報が付与され、『どの文献が何回アクセスされた』という情報として出力できることが強み。
- JISCのPIRUS (Publisher and institutional repository usage statistics, イギリス)
COUNTER準拠の個別記事レベルの標準規格と利用レポートを策定すること。
JISC, COUNTER, リポジトリ, 出版社・ベンダーへの提案を含む最終レポートを報告している。(2009,1, ROATで翻訳済)

COUNTER準拠にする意味

雑誌の出版社統計と同じ基準を使用することによって、雑誌の統計とセルフアーカイブのダウンロード数を足して、特定論文について、全部で何回利用されたかが測定できるようになるのです。

- DINIのOAS (Open-Access-Statistik, ドイツ) 
アクセスログの分析を行うシステムの提供。
基本となる考え方は、文献がどれくらい利用されているか。
つまり **Usage Activity ≠ Usage Statistics**
論文ごとに複数の統計基準で分析できるようにしている。
(COUNTER, IFABC, LogEC)。



- Lille第3大学(フランス)のプロジェクトチーム
各リポジトリがそれぞれに公表している利用統計に対して、相互比較を可能にするための概念枠組みを提供しようとしている。
用語等の定義や、統計結果報告の様式標準化を目指す。